

PLENACOUSTIC FUNCTION ON THE CIRCLE WITH APPLICATION TO HRTF INTERPOLATION

Thibaut Ajdler*, Luciano Sbaiz* and Martin Vetterli*[◇]

*Laboratory for Audio-Visual Communications (LCAV)

Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

[◇]Department of EECS, University of California at Berkeley, Berkeley CA 94720, USA

Email: {thibaut.ajdler, luciano.sbaiz, martin.vetterli}@epfl.ch

ABSTRACT

The sampling of the soundfield using circular microphone arrays is studied. We give a new representation of the soundfield sampled on such an array using the two-dimensional Fourier transform. We then show how we can reconstruct the soundfield at all possible positions on the array from a finite number of measurements. We give the necessary azimuthal angular sampling frequency necessary to reconstruct the data up to some temporal frequency. This new technique is applied to interpolate Head Related Transfer Functions (HRTFs) where we also derive an angular sampling theorem.

1. INTRODUCTION

In this paper we study the plenacoustic function on a circle. This representation models the soundfield everywhere in space (e.g. in a room). The plenacoustic function has already been studied to sample the soundfield along linear or planar arrays in [1, 2, 3]. There, interesting results have been shown but interpolation is always suffering border effects due to the finite length of the array and good reconstruction results are only obtained in the middle of the array. The best results for interpolation are found using a circular array as will be shown in the present paper. Circular microphone and loudspeaker arrays are widely used in wave field synthesis and beam-forming systems [4]. A contribution of this paper defines exactly what angular sampling frequency needs to be used to reconstruct the soundfield up to some temporal frequency. The interpolation method to achieve reconstruction is also presented.

The theoretical results obtained in the first part of the paper are then applied to the interpolation of Head Related Transfer Functions (HRTFs). Much work has been done to achieve good interpolation of HRTFs (e.g. pole-zero approximation models [5], spherical spline-based methods [6]) but nevertheless the problem is still considered as an open question [7]. In this respect, application of our novel technique allows us to derive an angular sampling theorem for HRTFs. We define the necessary azimuthal angular spacing between HRTFs measurements in order to perfectly reconstruct HRTFs at any azimuth up to a specified temporal frequency. This can be done for all possible elevation angles. With this sampling criterium we obtain interesting interpolation results on both simulated and real measurement data sets.

The work presented in this paper was supported in part by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

2. PLENACOUSTIC FUNCTION ON THE CIRCLE

In order to construct the plenacoustic function (PAF) for a circular array, we need to consider the scheme shown in Fig. 1(a). The

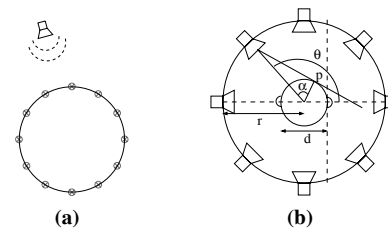


Fig. 1. Circular arrays. (a) Circular microphone array. (b) Typical setup to record someone's HRTFs.

source is located at position $S(s_x, s_y, s_z)$ and the different microphones are at positions $M(m_x, m_y, m_z)$. The PAF on the circle, is a continuous function that contains all the room impulse responses from the position of the source to all the possible microphones positions on the circle. In a first scenario we will present results in free field situation. Further using the image model method [8], we will show that the same results are obtained when one considers the complete room impulse response.

2.1. Construction of the plenacoustic function in free field

One can write the positions of the microphones in cylindrical coordinates as follows: $m_x = o_x + r \cos \theta$, $m_y = o_y + r \sin \theta$. Without loss of generality and for simplicity of the following calculations, consider a rotation of the axis such that the y component of the source is canceled. Further consider the center of coordinates to be the center of the circle formed by the microphones. Therefore, we have that $o_x = o_y = 0$. In free field, we have the following expression for the time traveled by the wave from the source to the microphones:

$$t = \frac{\sqrt{(s_x - r \cos \theta)^2 + (r \sin \theta)^2 + (s_z - m_z)^2}}{c} \quad (1)$$

with c the speed of sound propagation. The function given in (1) has a shape that can be seen in Fig. 2(a). The time taken by the sound wave to arrive at each microphone is given as function of the angle θ . Fig. 2(b) represents the Fig. 2(a) as seen from above.

For simplicity, this representation will be used for the rest of the paper. Note that in a first step we do not consider yet the attenuation depending on the distance traveled. It will be considered later in Section 2.3. We call q_a the PAF without attenuation and p_a the exact PAF.

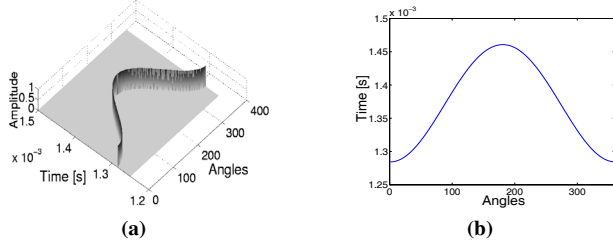


Fig. 2. Plenacoustic function on the circle. (a) PAF in 3D. (b) Top view of the PAF.

2.2. Plenacoustic sampling

The PAF as described above is a function of time and angle, $\tilde{q}_a(\theta, t)$. Note that the symbol $\tilde{\cdot}$ indicates a function defined on a continuous time and angle domain. We can study this function in the frequency domain. We need to take the 2-dimensional Fourier transform (2D-FT) of the gathered data by introducing the temporal pulsation Ω_t and the spatial pulsation L_θ . The obtained spectrum is $\tilde{Q}_a(L_\theta, \Omega_t)$. Note that the PAF on a circle is a periodic function of period 2π in θ and continuous in t . Therefore, L_θ only takes discrete values being $z\frac{2\pi}{T}$, with $z \in \mathbb{Z}$ and T being the period of 2π . L_θ is thus only defined for the integer values. In order to sample the PAF, we need to sample the room impulse responses at a certain temporal sampling pulsation η_t depending on the desired audio bandwidth. Further, by taking an evenly spaced finite number of impulse responses, we uniformly sample the PAF along the circle. We introduce $\eta_\theta = \frac{2\pi}{\Delta\theta}$ with $\Delta\theta$ the angular spacing between two consecutive microphones positions. When the PAF is sampled, repetitions of the spectrum occur. The discrete spectrum $Q_a(l_\theta, \omega_t)$ is given by the following expression:

$$Q_a(l_\theta, \omega_t) = \frac{1}{\Delta\theta\Delta t} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \tilde{Q}_a(l_\theta - \frac{2\pi k_1}{\Delta\theta}, \omega_t - \frac{2\pi k_2}{\Delta t}).$$

l_θ and ω_t are the angular pulsation and the temporal pulsation of the sampled PAF respectively. Note that taking the 2D-FT of the sampled PAF corresponds to taking a discrete Fourier transform in the time axis and a discrete Fourier series on the spatial axis, with $l_\theta \in \{0, 1, \dots, (\frac{2\pi}{\Delta\theta} - 1)\}$. The spectrum (without spectral repetitions) is shown in Fig. 3. We observe that most of the energy

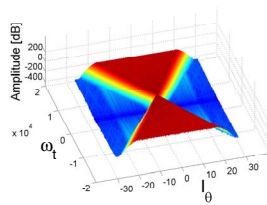


Fig. 3. 2D-FT of the circularly sampled room impulse responses. The spectrum is contained in a bow-tie region. Outside of that region, the energy is almost zero.

As we know that the 2D-FT of a line of slope m is another line of slope $-\frac{1}{m}$, we look for the maximal slope in (1). The obtained value is a lower bound for the support of the bow-tie region. Since the function is not a simple line, there is still some energy outside of the bow-tie region. However numerical computations show that the spatial decay of the spectrum appears to be exponential. This decay is dependent on the regularity of (1) and is still under investigation. The first derivative of (1) with respect to θ is given by

$$\frac{dt}{d\theta} = \frac{s_x r \sin \theta}{c \sqrt{(s_x - r \cos \theta)^2 + (r \sin \theta)^2 + (s_z - m_z)^2}} \quad (2)$$

To know where the maxima of this function occur, we calculate the second derivative of (1) with respect to θ . By setting the obtained expression to zero, we obtain the values of the angle θ that maximize the first derivative. Replacing these values in (2) leads us to the maximal value of the first derivative or to the maximal slope of (1):

$$\frac{dt}{d\theta} = \pm \frac{\sqrt{A + 2s_x r} - \sqrt{A - 2s_x r}}{2c} \quad (3)$$

with $A = s_x^2 + r^2 + (s_z - m_z)^2$. When the source is located in the same plane as the circular array, the expression of the slope gets simpler: for a source located inside of the circular array, the maximal slope of (1) is $\frac{dt}{d\theta} = \pm \frac{(s_x)}{c}$; for a source located outside of the circular array, the slope of (1) becomes $\frac{dt}{d\theta} = \pm \frac{(r)}{c}$. The maximal slope is thus only dependent on the minimum between the distance from the center of the array to the source and the radius of the array,

$$\frac{dt}{d\theta} = \pm \frac{(\min(s_x, r))}{c}. \quad (4)$$

Further we can also prove [3] that when the source is not in the plane of the circular array, the slope is always smaller or equal to (4).

$$\frac{dt}{d\theta} = \left| \frac{\sqrt{A + 2s_x r} - \sqrt{A - 2s_x r}}{2c} \right| \leq \left| \frac{\min(s_x, r)}{c} \right|.$$

This is an interesting result since all the reflections present in a real room corresponding to the virtual sources obtained by the image source method [8] contain different z -components than the the circular array. Therefore, these virtual sources lead to smaller slopes and therefore evolve slower in space. This allows us to conclude that the maximal slope in a real room (with reflections) is simply $\frac{r}{c}$. The slope of the triangular spectrum is now:

$$\frac{2c}{\sqrt{A + 2s_x r} - \sqrt{A - 2s_x r}} \geq \frac{c}{\min(s_x, r)}.$$

The minimal slope that we can observe in the bow-tie shape (in presence of reflections) in a typical room is therefore

$$\omega_t = \pm \frac{cl_\theta}{r}. \quad (5)$$

2.3. Effect of distance traveled

As one knows from the solution of the wave equation [9], the pressure decays proportionally to the distance traveled. We can express the real PAF as the product in time domain of the constant

amplitude PAF shown in Section 2.2 with a function that adds the amplitude variation. This can be written as

$$\tilde{p}_a(\theta, t) = a(\theta)\tilde{q}_a(\theta, t).$$

In frequency domain, it corresponds to the following convolution

$$\tilde{P}_a(L_\theta, \Omega_t) = (A(L_\theta)\delta(\Omega_t)) * \tilde{Q}_a(L_\theta, \Omega_t).$$

Because of the smooth character of $a(\theta)$, its Fourier transform is a very fast decaying function. The biggest variation in distance is obtained when the loudspeaker is in the plane of the microphones. We numerically show some $A(L_\theta)$ in Fig. 4.

$$A(L_\theta) = \int_{\theta=0}^{2\pi} \frac{e^{-jL_\theta\theta}}{r\sqrt{\rho^2 - 2\rho\cos\theta + 1}} d\theta \quad (6)$$

with $\rho = \frac{L_\theta}{r}$. For $L_\theta = 5$, we already have a very large decay of more than 60dB. Therefore, we consider that the effect of the convolution with $A(L_\theta)$ is very small and can be considered as negligible.

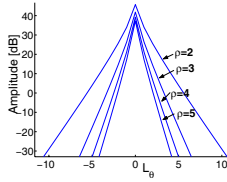


Fig. 4. $A(L_\theta)$ for different ρ .

2.4. Interpolation results

The interpolation is done using a usual sinc interpolator. This interpolation is very suitable in the case of a circular array since the Fourier transform is applied on a circular array that is 2π periodic. In the case of a linear array [2], interpolation performance decreases due to the finite length of the array. The spectrum of the PAF of an infinite linear array is decaying very fast [2], but once the PAF is windowed along the spatial axis, the spatial decay of its spectrum is affected.

We simulate room impulse responses on a circular array of radius 1 meter every 1° . This angle corresponds to the critically sampled angular spacing for a temporal sampling frequency of about 19 kHz. We interpolate the data by a factor 2 and we compare them with the corresponding simulations. We use a normalized mean squared error (MSE) criterium to evaluate the interpolation error. The MSE is zero on every odd positions (since these were the data used for the interpolation) and the error on the even spatial positions is of the order of -60 dB as can be seen in Fig. 5. These results indicate that the interpolation formula is quite accurate.

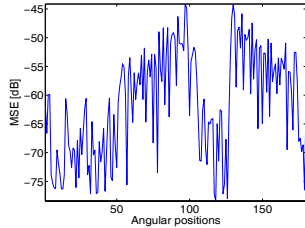


Fig. 5. MSE for the interpolated spatial positions.

3. HRTF INTERPOLATION

In Section 2, we have presented theoretical results for the sampling of the soundfield using a circular microphone array. The exact same theory can be applied for the dual problem, the case when we listen at one position the soundfield generated by a circular loudspeaker array. A very interesting application of this dual problem is the sampling of HRTFs in an anechoic chamber to measure the characteristics of the pinnae, head and torso of a person [10]. The typical setup for an HRTF measurement is shown in Fig. 1(b). The loudspeakers are located along a circle around the person. The microphone is located in the ear of the listener to capture the soundfield at the entrance of the ear. One can construct a "plen-HRTF" function by placing all the different head related impulse responses next to each other and take the 2-dimensional Fourier transform of this data. To understand the shape of the spectrum, we apply the same theory as the one presented in Section 2. When the head is well centered in the middle of the loudspeaker array, we consider the position of the microphone to be of the order of 9 cm (half the spacing between the two ears)[10]. We therefore can predict that in order to sample HRTFs, we have to satisfy the following relation:

$$\omega_t = \pm \frac{cl_\theta}{.09}. \quad (7)$$

This equation tells us that in order to sample HRTFs for a normal adult human being at 44.1 kHz, a spacing of 4.9° is necessary.

3.1. Head shadowing

The theory expressed above is valid in the case of HRTFs when one does not consider the effect of the head shadowing. In practice, the wave is diffracted by the head [10]. This diffraction has to be taken in account and modifies (1). We give first an intuition about the effect of the diffraction followed by a more accurate model. Consider different loudspeakers located at radius r and angle θ as shown in Fig. 1(b) (the following formulas are for $\theta < \pi$, similar formulas exist for $\theta > \pi$). For each loudspeaker position, consider the tangent to the head at point p . The point p has a radius of $\frac{d}{2}$ and an angle of $\theta - \alpha$, with $\cos \alpha = \frac{d/2}{r}$. For all the directions of $\theta < \alpha$, the time of arrival from the loudspeaker position follows (1). When $\theta \geq \alpha$, the time of arrival consists of the distance from the loudspeaker to the tangent point p , followed by the circle bow starting from p to the ear position. The time of arrival follows the expression:

$$t = \frac{\frac{d}{2}(\theta - \alpha) + \sqrt{r^2 - \frac{d^2}{4}}}{c}. \quad (8)$$

The two expressions (1) and (8) are shown in Fig. 6(a). By looking at the first derivative of the obtained waveform, one can observe in Fig. 6(b) that the slope does not increase when considering the head shadowing. It stays constant for angles $\theta \geq \alpha$. Therefore, considering the head shadowing does not modify (7).

This model is nevertheless unsatisfactory because of the cusp in Fig. 6(a) due to the simplistic diffraction model. A more accurate model is given by [11]. The HRTFs are expressed as:

$$H(\rho, \mu, \theta) = -\frac{\rho}{\mu} e^{-i\mu\rho\Psi} \quad (9)$$

with $\Psi(\rho, \mu, \theta) = \sum_{m=0}^{\infty} (2m+1) P_m(\cos \theta) \frac{h_m(\mu \rho)}{h'_m(\mu)}$, μ a normalized temporal frequency, P_m the Legendre polynomial of degree m and h_m the m^{th} order spherical Hankel function. Taking the discrete Fourier series of (9) we also observe a bow-tie spectrum satisfying (7), as shown in Fig. 7(a). The obtained spectrum of measured HRTFs [12] sampled every 5° in an anechoic chamber is shown in Fig. 7(b). We can observe that (7) is satisfied with a spacing of 5° almost no aliasing is noticed at CD sampling frequency (44.1 kHz).

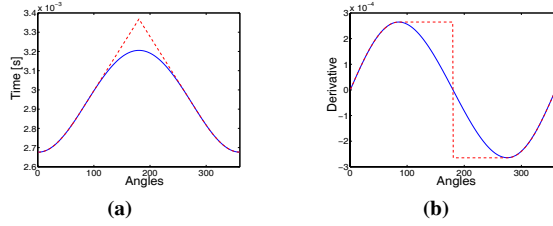


Fig. 6. (a) Time of arrival with (in full lines) or without (in dashed lines) shadowing in function of angle of the loudspeaker. (b) First derivative of the time of arrival with (in full lines) or without (in dashed lines) shadowing.

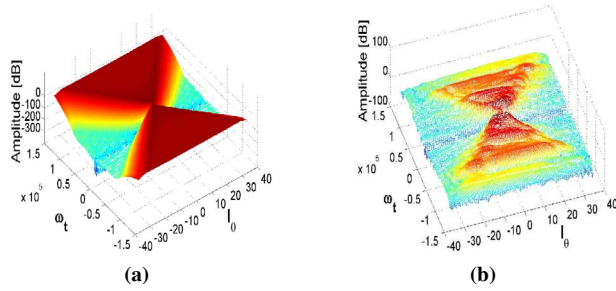


Fig. 7. 2D Spectra of HRTF (a) using a diffraction model. (b) using measured data.

3.2. Interpolation results

We have applied the same technique as presented in Section 2.4 for different data sets. We used 36 measurements spaced by 10° to interpolate the HRTFs by a factor 2. The interpolated measurements were compared with the original 72 measurements and the MSE was calculated. Using a spacing of 10° , we can interpolate the HRTFs up to 10.8 kHz as follows from (7). The first data set was a simple simulation in free field with a set of loudspeakers and a microphone copying the real setup of HRTFs measurements without considering the head shadowing. The results of interpolation are shown in full lines in Fig. 8(a). The same simulation was carried out using the model of HRTFs from [11] and the results are shown in dotted lines in Fig. 8(a). We see that these two experiments show a MSE varying from -35 to -65 dB. Finally using the experimental HRTFs from [12], the MSE was slightly higher but still of the order of -40 dB as can be seen in Fig. 8(b).

4. CONCLUSION

In this paper, we have studied the sampling of soundfields using circular microphone arrays. We have shown that the 2D-FT associated to the soundfield has a spectrum that has a bow-tie shape.

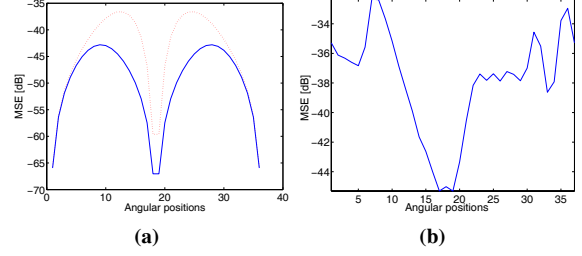


Fig. 8. Interpolation MSE. (a) Comparison of MSE for interpolation of simulated HRTFs with (in dotted lines) or without head shadowing (in full lines). (b) MSE on the interpolation of the measured HRTFs.

We have explained how reconstruction of the soundfield at all possible positions on the array is possible using a finite number of measurements. The proposed technique has been applied with success to the problem of interpolation of HRTFs where we computed the necessary azimuthal angular sampling frequency necessary to reconstruct the HRTFs up to some temporal frequency.

5. ACKNOWLEDGEMENT

We would like to thank C. Faller for his interesting remarks and comments.

6. REFERENCES

- [1] T. Ajdler and M. Vetterli, "The plenacoustic function and its sampling," in *MPCA workshop*, 2002.
- [2] T. Ajdler and M. Vetterli, "The plenacoustic function, sampling and reconstruction," in *IEEE ICASSP*, 2003.
- [3] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *Submitted to IEEE Trans. on Signal Proc.*, 2005.
- [4] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sound fields by wave field synthesis," in *110th Conv. of the AES*, 2001.
- [5] M.A. Blommer and G.H. Wakefield, "Pole-zero approximations for HRTF using a logarithmic error criterium," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 278–287, 1997.
- [6] Carlile et al., "Continuous virtual auditory space using HRTF interpolation: acoustic and psychophysical errors," in *Int. Symposium on Multimedia Information Processing*, 2000.
- [7] D. Ramani, D. N. Zotkin, and N.A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *IEEE ICASSP*, 2004.
- [8] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, 1979.
- [9] P.M. Morse and K.U. Ingard, *Theoretical Acoustics*, McGraw-Hill, 1968.
- [10] J. Blauert, *Spatial hearing*, MIT press, 2001.
- [11] R. O. Duda and W.M. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, pp. 3048–3058, 1998.
- [12] B. Cardner and K. Martin, "HRTF measurements of a Kemar dummy-head microphone," Tech. Rep., MIT Media Lab, 1994.